Economics Letters 120 (2013) 437–441

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On a general class of long run variance estimators[☆☆]

Xianyang Zhang[*], Xiaofeng Shao

University of Missouri-Columbia, United States
University of Illinois at Urbana-Champaign, United States

HIGHLIGHTS

- We propose a general class of LRV estimators in the GMM framework.
- The LRV estimator includes some recently developed estimators as special cases.
- First order asymptotics of the Wald statistics based on general LRV estimators.

ARTICLE INFO

Article history:
Received 12 April 2013
Accepted 30 May 2013
Available online 10 June 2013

JEL classification:
C12
C13

Keywords:
Fixed-smoothing asymptotics
Generalized method of moments
Long run variance matrix

ABSTRACT

This note proposes a class of estimators for estimating the asymptotic covariance matrix of the generalized method of moments (GMM) estimator in the stationary time series models. The proposed estimator is general enough to include the traditional heteroskedasticity and autocorrelation consistent (HAC) covariance estimator and some recently developed estimators, such as the cluster covariance estimator and projection-based covariance estimator, as special cases. We also study the first order asymptotics of the Wald statistics based on the general covariance estimators when the underlying smoothing parameter is held fixed.

1. Introduction

In stationary time series models, the asymptotic covariance matrix of the generalized method of moments (GMM) estimator is usually estimated nonparametrically by the kernel-based methods, where the bandwidth parameter is assumed to grow slowly with the sample size in the asymptotics (see Newey and West, 1987; Andrews, 1991). Recent studies on heteroskedasticity and autocorrelation consistent (HAC) based robust inference have developed alternative first order asymptotic theory (as compared to the traditional χ^2-based approximation), which was shown to provide more accurate approximation to the sampling distributions of the associated test statistics. For example, Kiefer and Vogelsang (2005, KV, hereafter) developed a first order asymptotic theory where the proportion of the bandwidth involved in the HAC estimator to the sample size T, denoted as b, is held fixed in the asymptotics. Using the higher-order Edgeworth expansions, Jansson (2004), Sun et al. (2008), Sun (2010) and Zhang and Shao (forthcoming) rigorously proved that the fixed-b asymptotics provides a high order refinement over the traditional small-b asymptotics in the Gaussian location model. Sun (2013) developed a procedure for hypothesis testing in time series models by using the nonparametric series method. The basic idea is to project the time series onto a space spanned by a set of Fourier basis functions (see Phillips, 2005, for an early development) and construct the covariance matrix estimator based on the projection vectors with the number of basis functions held fixed. Also see Sun (2011) for the use of a similar idea in the inference of the trend regression models. Ibragimov and Müller (2010) proposed a subsampling based t-statistic for robust inference where the unknown dependence structure can be in the temporal, spatial or other forms. In their paper, the number of non-overlapping blocks is held fixed. The t-statistic based approach was extended by Bester et al. (2011) to the inference of spatial and panel data with group structure. In the context of misspecification testing, Chen and Qu (forthcoming) proposed a modified M test of Kuan and Lee (2006) which

[☆☆] This note is drawn from an early working paper entitled "Fixed-smoothing asymptotics for time series" (arXiv:1204.4228). A substantial part of the working paper appeared in Zhang and Shao (forthcoming), which has little overlap with this note.

[*] Correspondence to: Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, United States. Tel.: +1 573 772 1577; fax: +1 573 884 5524.
E-mail address: zhangxianyang@gmail.com (X. Zhang).

0165-1765 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-SA license.
http://dx.doi.org/10.1016/j.econlet.2013.05.026

involves dividing the full sample into several recursive subsamples and constructing a normalization matrix based on them. In the statistical literature, Shao (2010) developed the self-normalized approach to inference for time series data that uses an inconsistent long run variance (LRV) estimator based on recursive subsample estimates. The self-normalized method is an extension of Lobato (2001) from the sample autocovariances to more general approximately linear statistics and it coincides with KV's fixed-b approach in the inference of the mean of a stationary time series by using the Bartlett kernel and letting $b = 1$. Although the above inference procedures are proposed in different settings and for different problems and data structure, they share a common feature in the sense that the underlying smoothing parameters in the asymptotic covariance matrix estimator such as the number of basis functions, the number of cluster groups and the number of recursive subsamples, play a similar role as the bandwidth in the HAC estimator.

The goal of this note is to introduce a general class of estimators for estimating the LRV matrix in the inference of stationary time series models estimated by GMM. Our proposal includes the traditional lag window type (or HAC) covariance estimator, the projection-based covariance estimator, the cluster-based covariance estimator and the blockwise recursive subsampling-based covariance estimator as special cases. The general covariance estimator considered here involves projecting the original data onto a space spanned by a sequence of basis functions (not necessarily orthogonal), where the number of basis functions K plays a key role in determining asymptotic properties of the estimator. Under the fixed-K asymptotics, we show that the Wald statistic based on the general LRV estimator converges to an (approximate) F distribution with a scale constant depending only on K and the number of restrictions being tested. Thus our result provides a unification of the various recently proposed fixed-smoothing inference procedures in the first order sense.

We introduce some notation. Denote by $\lfloor a \rfloor$ the integer part of a real number a. Let $L^2[0, 1]$ be the space of square integrable functions on $[0, 1]$. Denote by $D[0, 1]$ the space of functions on $[0, 1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Denote by "$\Rightarrow$" weak convergence in the $\mathbb{R}^{q_0}$-valued function space $D^{q_0}[0, 1]$, where $q_0 \in \mathbb{N}$. Define "$\rightarrow^d$" convergence in distribution. We use "$\otimes$" to denote the Kronecker product in matrix algebra. The notation $N(\mu, \Sigma)$ is used to denote the multivariate normal distribution with mean μ and covariance Σ. Let χ^2_k be a random variable following χ^2 distribution with k degrees of freedom.

2. Basic setup and assumptions

In linear and nonlinear models with moment conditions, it is standard to employ GMM to estimate the model parameters. We follow the GMM setup as described in KV. Consider a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$ of interest, where Θ is the parameter space. Denote θ_0 the true parameter of θ which is an interior point of Θ. Let y_t denote a vector of observed data and assume the moment conditions

$$E[f(y_t, \theta)] = 0, \quad t = 1, 2, \ldots, T \tag{1}$$

hold if and only if $\theta = \theta_0$, where $f(\cdot)$ is $m \times 1$ vector of functions with $m \geq d$ and $\operatorname{rank}(E[\partial f(y_t, \theta_0)/\partial\theta']) = d$. When $m > d$, the parameter θ is over-identified with the degree of over-identification $v = m - d$. Define the partial sum $g_t(\theta) = T^{-1}\sum_{j=1}^t f(y_j, \theta)$. Then the GMM estimator of θ_0 is given by

$$\hat{\theta}_T = \operatorname*{argmin}_{\theta\in\Theta} g_T(\theta)' W_T g_T(\theta), \tag{2}$$

where W_T is a $m \times m$ semi-positive definite weighting matrix. Further define

$$G_t(\theta) = (G_{t1}(\theta), \ldots, G_{tm}(\theta))' = \frac{\partial g_t(\theta)}{\partial\theta'} = \frac{1}{T}\sum_{j=1}^t \frac{\partial f(y_j, \theta)}{\partial\theta'}.$$

Using the mean value theorem for each element of g_T, we have $g_T(\hat{\theta}_T) = g_T(\theta_0) + \tilde{G}_T(\hat{\theta}_T - \theta_0)$, where $\tilde{G}_T = (G_{T1}(\tilde{\theta}_{T1}), \ldots, G_{Tm}(\tilde{\theta}_{Tm}))'$ and $\tilde{\theta}_{Tj}$ is between θ_0 and $\hat{\theta}_T$ for each $1 \leq j \leq m$. Note that $G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0$ by the first order condition, which implies that

$$G_T(\hat{\theta}_T)' W_T g_T(\theta_0) + G_T(\hat{\theta}_T)' W_T \tilde{G}_T(\hat{\theta}_T - \theta_0)$$
$$= G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0.$$

Solving the above equation, we have

$$T^{1/2}(\hat{\theta}_T - \theta_0) = -(G_T(\hat{\theta}_T)' W_T \tilde{G}_T)^{-1} G_T(\hat{\theta}_T)' W_T (T^{1/2} g_T(\theta_0)).$$

To derive the asymptotic distribution of $\hat{\theta}_T$, we make the following high-level assumptions as KV and Sun (2010).

Assumption 2.1. $\hat{\theta}_T \rightarrow^p \theta_0$.

Assumption 2.2. $T^{1/2} g_{\lfloor Tr \rfloor}(\theta_0) \Rightarrow \Delta W_m(r)$ where

$$\Delta\Delta' = \Omega = \sum_{j=-\infty}^{+\infty} E[f(y_t, \theta_0) f(y_{t-j}, \theta_0)'],$$

and $W_m(r)$ is a m—dimensional vector of independent standard Brownian motions.

Assumption 2.3. $\tilde{G}_T \rightarrow^p G_0$ uniformly for all $\tilde{\theta}_{Tj}$ between $\hat{\theta}_T$ and θ_0, where $G_0 = E[\partial f(y_j, \theta_0)/\partial\theta']$ and $1 \leq j \leq m$.

Assumption 2.4. The weighting matrix W_T is symmetric and semi-positive definite such that $W_T \rightarrow^p W_0$ and $G_0' W_0 G_0$ is positive definite.

Under Assumptions 2.1–2.4, it is easy to see that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \rightarrow^d -(G_0' W_0 G_0)^{-1} G_0' W_0 \Delta W_m(1) =^d N(0, V_0),$$

where "$=^d$" denotes "equal in distribution" and the asymptotic covariance matrix $V_0 := (G_0' W_0 G_0)^{-1} G_0' W_0 \Omega W_0 G_0 (G_0' W_0 G_0)^{-1}$. To make inference on θ_0, we have to estimate G_0, W_0 and the LRV matrix Ω. Under the above assumptions, G_0 and W_0 can be consistently estimated by their sample counterparts $G_T(\hat{\theta}_T)$ and W_T respectively. It remains to estimate the LRV matrix Ω. In the next section, we introduce a general class of estimators for Ω and V_0.

3. LRV estimators

To present the idea, we focus on the hypothesis testing problem that $H_0 : r(\theta_0) = 0$ versus the alternative that $H_a : r(\theta_0) \neq 0$, where $r(\theta)$ is a $p \times 1$ continuously differentiable function with the first order derivative matrix $R(\theta) = \partial r(\theta)/\partial\theta'$ and $p \leq d$. Let

$$\hat{V}_T = (G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1}$$
$$\times (G_T(\hat{\theta}_T)' W_T \hat{\Omega}_T W_T G_T(\hat{\theta}_T))(G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1},$$

be an estimator of V_0, where $\hat{\Omega}_T$ is the LRV estimate of Ω. The Wald statistic for testing H_0 against H_a is defined as

$$F_T = T r(\hat{\theta}_T)' \hat{D}_T^{-1} r(\hat{\theta}_T)/p, \tag{3}$$

where $\hat{D}_T = R(\hat{\theta}_T) \hat{V}_T R(\hat{\theta}_T)'$. The widely used lag window type LRV estimator is given by

$$\hat{\Omega}_T = \frac{1}{T}\sum_{i=1}^T \sum_{j=1}^T \mathcal{K}\left(\frac{i-j}{bT}\right) f(y_i, \hat{\theta}_T) f(y_j, \hat{\theta}_T)', \tag{4}$$

where $\mathcal{K}(\cdot)$ is a kernel function and b is the proportion of the truncation lag to the sample size. By setting

$$\hat{u}_i = R(\hat{\theta}_T)(G_T(\hat{\theta}_T)'W_TG_T(\hat{\theta}_T))^{-1}G_T(\hat{\theta}_T)'W_Tf(y_i, \hat{\theta}_T),$$

we have

$$\hat{D}_T = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathcal{K}\left(\frac{i-j}{bT}\right)\hat{u}_i\hat{u}_j'.$$

When $\mathcal{K}(\cdot)$ is semi-positive definite, by Mercer's theorem, we have the spectral decomposition,

$$\mathcal{K}(r-t) = \sum_{j=1}^{+\infty}\lambda_j\phi_j(r)\phi_j(t), \quad 0 \le r, t \le 1/b, \tag{5}$$

where $\{\lambda_j\}$ and $\{\phi_j\}$ are the eigenvalues and orthonormal eigenfunctions corresponding to the kernel function respectively. We thus have the representation,

$$\hat{D}_T = \sum_{s=1}^{K}\lambda_s\left\{\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\phi_s\left(\frac{i}{bT}\right)\hat{u}_i\right\}\left\{\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\phi_s\left(\frac{j}{bT}\right)\hat{u}_j'\right\},$$

with $K = +\infty$. In the traditional asymptotics, b goes to zero as T increases which is referred as the small-b asymptotics. When $b \in (0, 1]$ is held fixed, it corresponds to the fixed-b asymptotics in KV. As pointed out in some recent studies (see e.g., Bester et al., 2011; Sun, 2011, 2013; Chen and Qu, forthcoming), K can also be held as a fixed positive integer, which can lead to a more accurate first order approximation. In light of these recent findings, we introduce a general class of estimators to estimate the LRV matrix. With a slight abuse of notation, we let $\{\phi_s(t)\}_{s=1}^K$ be a sequence of linearly independent functions in $L^2[0, 1/b]$ and $\{\lambda_j\}$ be a sequence of nonnegative weights such that $\sum_{j=1}^K\lambda_j = 1$. A set of elements $\{\psi_i\}_{i=1}^K$ in a real valued vector space is called linearly independent if and only if $\sum_{i=1}^K a_i\psi_i = \mathbf{0} \Rightarrow a_i = 0$ for $i = 1, 2, \ldots, K$. Here $\mathbf{0}$ denotes the null element in the vector space. Note that λ_j's in (5) are nonnegative when we consider semi-positive definite kernels in (4). Further let $V_s = \frac{1}{\sqrt{T}}\sum_{i=1}^T\phi_s\left(\frac{i}{bT}\right)\hat{u}_i$, be the normalized inner product between $\{\hat{u}_i\}_{i=1}^T$ and $\{\phi_s(i/(bT))\}_{i=1}^T$. Define $R = (R_{ij})_{i,j=1}^K$ with $R_{ij} = \int_0^1\tilde{\phi}_i(t/b)\tilde{\phi}_j(t/b)dt$, where $\tilde{\phi}_s(t/b) = \phi_s(t/b) - \int_0^1\phi_s(t/b)dt$, and $L = (L_{ij})_{i,j=1}^K$ an upper triangular matrix based on the Cholesky decomposition of R^{-1}, i.e., $L'L = R^{-1}$. Define $V = (V_1', V_2', \ldots, V_K')'$ and

$$V^* = (V_1^{*'}, V_2^{*'}, \ldots, V_K^{*'})' = (L \otimes I_p)V,$$

where $V_i^* = \sum_{j=1}^K L_{ij}V_j$ for $1 \le i \le K$. Then the general LRV estimator is given by

$$\hat{D}_T = \sum_{s=1}^{K}\lambda_sV_s^*V_s^{*'} = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\left\{\sum_{s=1}^{K}\lambda_s\sum_{m=1}^{K}L_{sm}\phi_m\left(\frac{i}{bT}\right)\right. \\ \left.\times\sum_{l=1}^{K}L_{sl}\phi_l\left(\frac{j}{bT}\right)\right\}\hat{u}_i\hat{u}_j', \tag{6}$$

and the test statistic based on the general LRV estimator is defined as,

$$F_T = \left[\sqrt{T}r(\hat{\theta}_T)\right]'\hat{D}_T^{-1}\left[\sqrt{T}r(\hat{\theta}_T)\right]/p. \tag{7}$$

The matrix R is introduced for orthogonalization so that the limiting distribution of the test statistic F_T does not depend on the basis functions. Note that the choice of R is not unique (see Example 3.3). In what follows, we shall show that the recently developed nonparametric series covariance estimator (Sun, 2011, 2013), the recursive subsampling-based covariance estimator (Chen and Qu, forthcoming) and the cluster covariance estimator (CCE) (Bester et al., 2011) are all special cases of the general LRV estimator. Throughout Examples 3.1–3.3, we set $b = 1$ and $\lambda_j = 1/K$ for $j = 1, 2, \ldots, K$.

Example 3.1. Let $\{\phi_s(t)\}_{s=1}^K$ be a sequence of orthonormal basis functions with $\int_0^1\phi_s(t)dt = 0$. Then we have $R = I_{K\times K}$ and $\hat{D}_T = \frac{1}{K}\sum_{j=1}^K V_jV_j'$, where $V_j = \frac{1}{\sqrt{T}}\sum_{i=1}^T\phi_j(i/T)\hat{u}_i$. When $\phi_s(t) = \sqrt{2}\sin(2\pi st)\left(\text{or } \phi_s(t) = \sqrt{2}\cos(2\pi st)\right)$, $s = 1, 2, \ldots, K$, it is straightforward to see that the LRV estimator corresponds to the series estimator considered in Sun (2011, 2013). In this case, the LRV estimator involves projecting the data onto a set of orthonormal basis and using the sample variance of the projection vectors, namely $\hat{D}_T$.

Example 3.2. For any fixed K with $K \le T$, we consider the basis function $\phi_s(t) = \mathbf{I}\{0 < t \le s/(K+1)\}$, $s = 1, 2, \ldots, K$, where $\mathbf{I}$ denotes the indicator function. Simple calculation gives us $R_{ij} = \int_0^1\tilde{\phi}_i(t)\tilde{\phi}_j(t)dt = \min(i,j)/(K+1) - (ij)/(K+1)^2$, and $\hat{D}_T = \frac{1}{K}\sum_{s=1}^K V_s^*V_s^{*'}$, where

$$V_s^* = \sqrt{\frac{K+1}{T}}\left(\sqrt{\frac{s+1}{s}}\sum_{i=1}^{\left\lfloor\frac{Ts}{K+1}\right\rfloor}\hat{u}_i - \sqrt{\frac{s}{s+1}}\sum_{i=1}^{\left\lfloor\frac{T(s+1)}{K+1}\right\rfloor}\hat{u}_i\right),$$

with $s = 1, 2, \ldots, K$ and $V_{K+1} = 0$. Therefore, the general LRV estimator reduces to the recursive subsampling-based estimator in Chen and Qu (forthcoming), where the idea is to divide the full sample into $K+1$ recursive subsamples and construct a normalization matrix based on the subsamples.

Example 3.3. Let $\{A_j\}_{j=1}^K$ be a partition of the unit intervals [0,1] with $K > p$. Suppose A_j is a finite union of disjoint intervals in [0,1]. Let $\phi_s(t) = \mathbf{I}(t \in A_s)$, $s = 1, 2, \ldots, K$. If we set $R_{ij} = \int_0^1\phi_i(t)\phi_j(t)dt$, then $L = \text{diag}\left(1/\sqrt{|A_1|}, 1/\sqrt{|A_2|}, \ldots, 1/\sqrt{|A_K|}\right)$, where $|A|$ denotes the Lebesgue measure of the set A. Further assume $|A_1| = |A_2| = \cdots = |A_K| = 1/K$, then we have

$$\hat{D}_T = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\sum_{s=1}^{K}\mathbf{I}(i/T \in A_s)\mathbf{I}(j/T \in A_s)\hat{u}_i\hat{u}_j' \\ = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathbf{I}(i,j \in \text{ the same group})\hat{u}_i\hat{u}_j',$$

where i is in group s if and only if $i/T \in A_s$, $s = 1, 2, \ldots, K$. In this case, the general LRV estimator is the same as the CCE considered in Bester et al. (2011), where the idea is to utilize the group structure in the observations and construct a covariance estimator based on the parameter estimates in each group. Using similar arguments in Sun (2010), we can show that

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor Tr\rfloor}\hat{u}_i \Rightarrow \Lambda B_p(r),$$

where Λ is an invertible matrix such that

$$\Lambda\Lambda' = R(\theta_0)(G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}R'(\theta_0)$$

and $B_p(r)$ denotes a p-dimensional vector of independent Brownian bridges. It implies that

$$\frac{1}{\sqrt{T}}\sum_{i\in s\text{th group}}\hat{u}_i \to^d \Lambda\int_{A_s} dB_p(r) =^d \frac{1}{\sqrt{K}}\Lambda(Z_s-\bar{Z}),$$

and

$$\hat{D}_T \to^d \frac{1}{K}\Lambda\sum_{s=1}^{K}(Z_s-\bar{Z})(Z_s-\bar{Z})'\Lambda',$$

where $(Z_1', Z_2', \ldots, Z_K')' \sim N(0, I_K \otimes I_p)$ and $\bar{Z} = \sum_{s=1}^{K} Z_s/K$. When $p = 1$, it is well known that

$$\sum_{s=1}^{K}(Z_s-\bar{Z})^2 =^d \chi^2_{K-1},$$

which implies $\sqrt{F_T} \to^d \sqrt{\frac{K}{K-1}}|t_{K-1}|$ under H_0. Note that $\sqrt{\frac{K-1}{K}F_T}$ coincides with the subsampling-based t-statistic in Ibragimov and Müller (2010) when we consider a location model and $r(\theta_0) = \theta_0 - \theta^*$ for a specific value θ^*. When $p > 1$, we have $F_T \to^d \frac{K}{K-p}F_{p,K-p}$. It is worth noting that the choice of $R = (R_{ij})$ with $R_{ij} = \int_0^1 \tilde{\phi}_i(t)\tilde{\phi}_j(t)dt$ is also valid. In this case, the limiting distribution of F_T would be a scaled F distribution with p numerator and $K-p+1$ denominator degrees of freedom (see Proposition 4.1).

Remark 3.1. For the subsampling-based inference, Assumption 2.2 can be relaxed by the assumptions which guarantee the finite dimensional convergence of $\left(\frac{1}{\sqrt{|\mathcal{G}_1|}}\sum_{i\in\mathcal{G}_1}\hat{u}_i, \ldots, \frac{1}{\sqrt{|\mathcal{G}_K|}}\sum_{i\in\mathcal{G}_K}\hat{u}_i\right)$. Here $\mathcal{G}_i$ is the set index for the ith group and $|\cdot|$ denotes the cardinality. When heteroscedasticity is present across different groups, the t-statistic tends to be conservative (see Ibragimov and Müller, 2010).

4. First order fixed-smoothing asymptotics

In what follows, we consider the first order fixed-smoothing asymptotics of the test statistic F_T based on the general LRV estimator under the null hypothesis and local alternatives. To emphasize the dependence on the smoothing parameter K, we shall use the notation $F_T(K)$ instead of F_T.

Proposition 4.1. *Suppose $p \le K < \infty$ and $b \in (0, 1]$ are both fixed. Let $R = (R_{ij})_{i,j=1}^K$ with $R_{ij} = \int_0^1 \tilde{\phi}_i(t/b)\tilde{\phi}_j(t/b)dt$ in the general LRV estimator. Further assume that $\phi_j(t)$ is continuously differentiable almost everywhere for $j = 1, 2, \ldots, K$. Under Assumptions 2.1–2.4 and H_0, we have*

$$F_T(K) \to^d Q_{p,K} := U_p' D_p^{-1} U_p/p, \tag{8}$$

where $D_p = \sum_{j=1}^K \lambda_j\eta_j\eta_j'$, $\{\eta_j\}_{j=1}^K$ and U_p are independent and identically distributed (iid) as $N(0, I_p)$. In particular, if $\lambda_j = 1/K$ for $j = 1, 2, \ldots, K$, we get

$$F_T(K) \to^d \frac{K}{K-p+1}F_{p,K-p+1}. \tag{9}$$

Remark 4.1. When the weights λ_j's are not equal and $p = 1$, D_p is a weighted sum of independent χ^2_1 random variables. The limiting null distribution $Q_{p,K}$ can be further approximated by a scaled F distribution with the parameters chosen properly to match the first two moments (see Sun, 2010). Compared to Sun (2013), we do not make the assumption that $\int_0^1 \phi_i(t)dt = 0$ and we allow the basis functions to be non-orthonormal (see Example 3.2). It is also worth noting that the above results hold when $\phi_s(t) = \mathbf{I}(t \in A_s)$ with A_s being a finite union of disjoint intervals in [0, 1].

Proposition 4.2. *Consider the local alternatives $H_a' : r(\theta_0) = c/\sqrt{T}$ with $c \neq \mathbf{0} \in \mathbb{R}^p$. Under the same assumptions in Proposition 4.1 with $\lambda_j = 1/K$, we have*

$$F_T(K) \to^d \frac{K}{K-p+1}F_{p,K-p+1,c'(R(\theta_0)V_0R(\theta_0)')^{-1}c},$$

where $F_{a,b,\delta}$ denotes the noncentral F distribution with degrees of freedom a and b, and noncentral parameter δ.

The proposition shows that the test $F_T(K)$ has non-trivial power against the local alternatives of order $1/\sqrt{T}$ and it is seen to be consistent if $\|c\| \to +\infty$ as $T \to +\infty$.

Proof of Proposition 4.1. Define $S_t(\hat{\theta}_T) = \frac{1}{T}\sum_{i=1}^t \hat{u}_i$. Using the continuous mapping theorem, we can show that

$$\sqrt{T}S_{\lfloor Tr\rfloor}(\hat{\theta}_T) = \frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor Tr\rfloor}\hat{u}_i \Rightarrow \Lambda B_p(r) :=^d \Lambda(W_p(r) - rW_p(1)),$$

where Λ is invertible such that $\Lambda\Lambda' = R(\theta_0)(G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}R(\theta_0)'$ and $W_p(r)$ is a p-dimensional vector of independent Brownian motions. Using summation by parts, we get

$$V_s = \frac{1}{bT}\sum_{t=1}^{T-1}\frac{[\phi_s\{t/(bT)\}] - \phi_s\{(t+1)/(bT)\}}{1/bT}\sqrt{T}S_t(\hat{\theta}_T) + \sqrt{T}\phi_s(1/b)S_T(\hat{\theta}_T),$$

where the last term disappears by recalling the fact that $G_T(\hat{\theta}_T)'W_Tg_T(\hat{\theta}_T) = 0$. By the continuous mapping theorem, we have

$$\begin{pmatrix} V_1 \\ \vdots \\ V_K \\ \sqrt{T}r(\hat{\theta}_T)\end{pmatrix} \to^d \begin{pmatrix} -\frac{\Lambda}{b}\int_0^1 \phi_1'(r/b)B_p(r)dr \\ \vdots \\ -\frac{\Lambda}{b}\int_0^1 \phi_K'(r/b)B_p(r)dr \\ \Lambda W_p(1)\end{pmatrix} =^d \begin{pmatrix} \Lambda\int_0^1 \tilde{\phi}_1(r/b)dW_p(r) \\ \vdots \\ \Lambda\int_0^1 \tilde{\phi}_K(r/b)dW_p(r) \\ \Lambda W_p(1)\end{pmatrix}.$$

Here we are using the fact that

$$-\frac{\Lambda}{b}\int_0^1 \phi_s'(r/b)B_p(r)dr = \Lambda\int_0^1 \phi_s(r/b)dB_p(r) = \Lambda\int_0^1\left\{\phi_s(r/b) - \int_0^1\phi_s(r/b)dr\right\}dW_p(r) = \Lambda\int_0^1 \tilde{\phi}_s(r/b)dW_p(r),$$

for $1 \le s \le K$. It is not hard to see that

$$\mathrm{Cov}\left(\int_0^1 \tilde{\phi}_s(r/b)dW_p(r), \int_0^1 dW_p(r)\right) = 0$$

and

$$\mathrm{Cov}\left(\int_0^1 \tilde{\phi}_s(r/b)dW_p(r), \int_0^1 \tilde{\phi}_t(r/b)dW_p(r)\right) = R_{st}I_p,$$

for $1 \le s, t \le K$, which implies

$$V = \left(V_1', V_2', \ldots, V_K', \sqrt{T}r(\hat{\theta}_T)'\right)' \to^d N(0, \tilde{R} \otimes \Lambda\Lambda'),$$

where $\tilde{R} = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}$.

We thus get $V^* = (L \otimes I_p)V \to^d N(0, LRL' \otimes \Lambda\Lambda') =^d N(0, I_K \otimes \Lambda\Lambda')$. In other words, V^* is free of the effect of the basis functions asymptotically. Recall that $\hat{D}_T = \sum_{s=1}^K \lambda_s V_s^* V_s^{*'}$, it is not hard to see that

$$F_T(K) = \left(\Lambda^{-1}\sqrt{T}r(\hat{\theta}_T)\right)' \{\Lambda^{-1}\hat{D}_T(\Lambda^{-1})'\}^{-1} \times \left(\Lambda^{-1}\sqrt{T}r(\hat{\theta}_T)\right)/p \to^d U_p' D_p^{-1} U_p/p,$$

where $D_p = \sum_{j=1}^K \lambda_j \eta_j \eta_j'$ and $\{\eta_j\}_{j=1}^K$ and U_p are iid with distribution $N(0, I_p)$. When $\lambda_j = 1/K, j = 1, 2, \ldots, K$, it is straightforward to see that $F_T(K) \to^d \frac{K}{K-p+1} F_{p,K-p+1}$. □

Proof of Proposition 4.2. Notice that $\sqrt{T}r(\hat{\theta}_T) \to^d N(c, \Lambda\Lambda')$ under the local alternatives. The result follows from the arguments in the proof of Proposition 4.1 and Theorem 5.2.2 in Anderson (2003). □

Acknowledgment

Shao's research is supported in part by National Science Foundation grant DMS-1104545.

References

Anderson, T.W., 2003. An Introduction to Multivariate Statistical Analysis. Wiley, New York.
Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.
Bester, C.A., Conley, T.G., Hansen, C.B., 2011. Inference with dependent data using cluster covariance estimators. Journal of Econometrics 165, 137–151.
Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley, New York.
Chen, Y., Qu, Z., 2012. *M* tests with a new normalization matrix. Econometric Reviews (forthcoming).
Ibragimov, R., Müller, U.K., 2010. *t*-statistic based correlation and heterogeneity robust inference. Journal of Business and Economic Statistics 28, 453–468.
Jansson, M., 2004. On the error of rejection probability in simple autocorrelation robust tests. Econometrica 72, 937–946.
Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory 21, 1130–1164.
Kuan, C.M., Lee, W.M., 2006. Robust *M* tests without consistent estimation of the asymptotic covariance matrix. Journal of the American Statistical Association 101, 1264–1275.
Lobato, I.N., 2001. Testing that a dependent process is uncorrelated. Journal of the American Statistical Association 96, 1066–1076.
Newey, W.K., West, K.D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Phillips, P.C.B., 2005. HAC estimation by automated regression. Econometric Theory 21, 116–142.
Shao, X., 2010. A self-normalized approach to confidence interval construction in time series. Journal of the Royal Statistical Society, Series, B. 72, 343–366.
Sun, Y., 2010. Let's fix it: fixed-*b* asymptotics versus small-*b* asymptotics in heteroscedasticity and autocorrelation robust inference. Working paper, Department of Economics, UCSD.
Sun, Y., 2011. Robust trend inference with series variance estimator and testing-optimal smoothing parameter. Journal of Econometrics 164, 345–366.
Sun, Y., 2013. A Heteroskedasticity and autocorrelation robust F Test using an orthonormal series variance estimator. The Econometrics Journal 16, 1–26.
Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroscedasicity-autocorrlation robust testing. Econometrica 76, 175–194.
Zhang, X., Shao, X., 2013. Fixed-smoothing asymptotics for time series. Annals of Statistics (forthcoming).